

AC P09799;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE VICTILIN GCT7-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
CC SEQUENCE FROM N.A.
RA Chian C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cotyledon seed embryogenesis and alpha
germination. XIX. Sequences and genomic organization of the alpha
globulin (vicilin) genes of cottonseed.";
RL Plant Mol. Biol. 9:533-546(1987).
CC !- FUNCTION: SEED STORAGE PROTEIN
CC !- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC !- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CONVICILIN, CONGLOMGIN, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to licensee@sb-sib.ch).
CC -----
EMBL; M19378; AAA30169.1; --.
PIR; S06398; S06398.
HSSP; 150477; ICAX.
INTERPRO; IPR001113; --.
DR PFAM: PF00546; Seedstorage_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605
SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 50.9%; Score 140; DB 1; Length 605;
Best Local Similarity 54.5%; Pred. No. 4.1e-08;
Matches 24; Conservative 11; Mismatches 7; Indels 2; Gaps 1;
OY :||:||||:|||:|||:|||:|||:|||:
Db 34 EDDPCQRYEDCRKRCQLETRGQTQPKCDRSEQQL-KEEQQR 75

RESULT 3
AGR_LUFCY STANDARD; PRT; 47 AA.
ID AGR_LUFCY
AC P56568;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DR 6.5k KDA ARGinine/Glutamate Rich Polypeptide (6.5k-AGRP).
OS Luffa cylindrica (Smooth Loofah) (Sponge gourd).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Luffa.
RN [1]
CC SEQUENCE.
RC TISSUE=SEED;
RX MEDLINE:97357433; PUBMED=9214759;
RA Kimura M., Park S.S., Sakai R., Yamasaki N., Funatsu G.;
RT "Primary structure of 6.5-k-arginine/glutamate-rich polypeptide from
the seeds of sponge gourd (luffa cylindrica)." ;
RL Biosci. Biotechnol. Biochem. 61:984-988 (1997).
CC !- FUNCTION: STORAGE PROTEIN WHICH PROVIDES NITROGEN AND CARBON
RESERVES DURING GERMINATION AND SEEDLING GROWTH.
CC !- MASS SPECTROMETRY: MW=5693.39; METHOD=MALDI.

RESULT 4
VSWA_TRYBR STANDARD; PRT; 471 AA.
ID VSWA_TRYBR
AC P20346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-MAY-1992 (Rel. 22, Last annotation update)
DE VARIANT SURFACE GLYCOPROTEIN WRATATA PRECURSOR (VSG).
OS Trypanosoma brucei rhodesiense.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatida; Trypanosoma.
RN [1]
CC SEQUENCE FROM N.A.
RC STRAIN=WRATAT 1;
RX MEDLINE=90290520; PubMed=2357229;
RA Reddy L.V., Hall T., Donelson J.E.;
RT "Sequences of three VSG mRNAs expressed in a mixed population of
Trypanosoma brucei rhodesiense.";
RL Biomed. Biophys. Res. Commun. 169:730-736(1990).
CC !- FUNCTION: VSG FORMS A COAT ON THE SURFACE OF THE PARASITE. THE
TYPANOSOME EVADES THE IMMUNE RESPONSE OF THE HOST BY EXPRESSING
A SERIES OF ANTIGENICALLY DISTINCT VSGS FROM AN ESTIMATED 1000
VSG GENES.
CC !- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR. A
A SOLUBLE FORM IS RELEASED FROM RUPTURED CELLS BY THE ACTION OF A
PI-PLC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to licensee@sb-sib.ch).
CC -----
EMBL; M33823; AAA30316.1; --.
DR PIR; A35480;
KW Glycoprotein; Antigen; Trypanosomiasis; GPI-anchor; Membrane; Signal.
FT SIGNAL 1 18
FT CHAIN 19 454
FT VARIANT SURFACE GLYCOPROTEIN WRATATA.
FT PROPERP 455 471 HYDROPHOBIC, REMOVED DURING MATURATION.
FT LIPTD 454 454 GPI-ANCHOR (BY SIMILARITY).
FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 449 449 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 471 AA; 50537 MW; 57ECCC08210423760 CRC64;

Query Match 24.5%; Score 67.5; DB 1; Length 471;
Best Local Similarity 31.0%; Pred. No. 1.8;
Matches 13; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

OY 3 DDPKRYEDCRRCEWDTRGOKEOOCEESCKS0YGEKDOQ 44
Db 402 NDKAKETE-CNSPOCKWDEKEKKRKLSEGRQAEKENQE 442

RESULT 5

MJDI_HUMAN
ID MJDI_HUMAN STANDARD; PRT; 360 AA.
AC P2425;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE MACHADO-JOSEPH DISEASE PROTEIN 1.
GN MJDI OR MJD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
RN [1]
RP SOURCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=95119166; PubMed=7874163;
RA Kawaguchi Y., Okamoto T., Tanivaki M., Aizawa M., Inoue M., Akiguchi I.,
Kawaguchi S., Nakamura H., Nakamura S., Nishimura M., Akiguchi I.,
Kimura J., Narumiya S., Kakizuka A.;
RT "CAG expansions in a novel gene for Machado-Joseph disease at
chromosome 14q32.1.";
RL Natl. Genet. 8:221-228(1994).
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF THE MACHADO-JOSEPH PROTEIN IS
HIGHLY POLYMORPHIC (14 TO 40 REPEATS) IN THE NORMAL POPULATION AND
IS EXPANDED TO ABOUT 68-82 REPEATS IN MJDI PATIENTS. LONGER
EXANSIONS RESULT IN EARLIER ONSET AND MORE SEVERE CLINICAL
MANIFESTATIONS OF THE DISEASE.
-!- DISEASE: DEFECTS IN MJDI ARE THE CAUSE OF MACHADO-JOSEPH DISEASE
(MJD). A NEURODEGENERATIVE DISORDER CHARACTERIZED BY CEREBELLAR
ATAXIA, PYRAMIDAL AND EXTRAPYRAMIDAL SIGNS, PERIPHERAL NERVE
PALSY, EXTERNAL OPHTHALMOPLEGIA, FACIAL AND LINGUAL FASCICULATION
AND BULGING. THIS DISEASE IS AUTOSOMAL AND DOMINANT, WITH A LATE
ONSET OF SYMPTOMS, GENERALLY AFTER THE FOURTH DECADE.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
in the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC MIM: 109150; -.
DR INTERPRO: IPR02950; -.
DR PFAM: PF02099; Josephin; 1.
DR PRINTS: PRO1233; JOSEPHIN.
KW Polymorphism; Triplet repeat expansion.
FT DOMAIN 292 317 poly-Gln.
SQ SEQUENCE 360 AA; 41531 MW; AD6774F20A6CD95D CRC64;

RESULT 6
GRPE_HELPY STANDARD; PRT; 189 AA.
ID GRPE_HELPY
AC P55970;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GRPE PROTEIN (HSP-70 COFACTOR).
GN GRPE OR HP010.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]

RP SOURCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97194467; PubMed=5252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Kline H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirchner E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Wadman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "the complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
CC -!- FUNCTION: STIMULATES, JOINTLY WITH DNAJ, THE ATPASE ACTIVITY OF
DNAK. HELPS TO RELEASE ADP FROM DNAK THUS ALLOWING DNAK TO RECYCLE
MORE EFFICIENTLY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GRPE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
in the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC MIM: 109150; -.
DR INTERPRO: IPR000740; -.
DR PFAM: PF01025; GRPE; 1.
DR PRINTS: PRO077; GRPEPROTEIN.
DR PROSITE: PS01071; GRPE; 1.
DR CHaperone; Heat shock.
SQ SEQUENCE 189 AA; 22040 MW; 62C0991B684FOC2D CRC64;

Query Match 22.5%; Score 62; DB 1; Length 360;
Best Local Similarity 35.0%; Pred. No. 5.5;
Matches 14; Conservative 12; Mismatches 12; Indels 2; Gaps 1;
Oy 10 EDCCRCE-WDTRQKEQQQCEESCKSQYGERDKDQQRHR 47
Db 279 EBLRKRRREAYFEKQQQKQQQQQQQQQQQQQQQQQQ 318
RESULT 7
TAGB_DICDI STANDARD; PRT; 1905 AA.
ID TAGB_DICDI
AC P54683;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PRESTALK-SPECIFIC PROTEIN TAGB PRECURSOR (EC 3.4.21.-).
GN TAGB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Dictyosteliida; Dictyostelium.
RN [1]
RP SOURCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=95262903; PubMed=7744252;
RA Shaulsky G., Kuspa A., Loomis W.F.;
RT "A multidrug resistance transporter/serine protease gene is required
for prestalk specialization in Dictyostelium.";
RL Genes Dev. 9:1111-1122(1995).
CC -!- FUNCTION: INTERCELLULAR COMMUNICATION VIA TAGB MAY MEDIATE
INTTEGRATION OF CELULAR DIFFERENTIATION WITH MORPHOGENESIS.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO PEPTIDASE FAMILY
CC S8; ALSO KNOWN AS THE SUBTILASE FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING
TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY.
CC -!- SIMILARITY: STRONG, TO TAGC.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; U20432; AAA62212.1; -
DR HSSP; P13569; INBD.
DR DICTYDB; DD02059; TAGB.
DR INTERPRO; IPR00205; -.
DR INTERPRO; IPR00110; -.
DR PFAM; PF00664; ABC_tran; 1.
DR PROSITE; PS00137; SUBTILISIN.
DR PRINTS; PR0073; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASp; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_His; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR KW hydrolase; Serine protease; ATP-binding; Transport; Transmembrane;
KW Signal.
FT SIGNAL 1 ?
FT CHAIN 2 ?
FT DOMAIN 378 ?
FT DOMAIN 1905 ?
FT TRANSMEM 1011 1031
FT TRANSMEM 1076 1096
FT TRANSMEM 1121 1141
FT TRANSMEM 1210 1230
FT TRANSMEM 1309 1329
FT TRANSMEM 1352 1387
FT ACT_SITE 432 432
FT ACT_SITE 695 695
FT NP_BIND 1553 1560
FT DOMAIN 63 67
FT DOMAIN 95 104
FT DOMAIN 107 134
FT DOMAIN 311 321
FT DOMAIN 833 839
FT DOMAIN 838 844
FT DOMAIN 871 876
FT DOMAIN 1012 1015
FT DOMAIN 1386 1389
FT DOMAIN 1398 1404
FT DOMAIN 1445 1450
FT DOMAIN 1765 1779
FT DOMAIN 1782 1785
FT DOMAIN 1807 1812
FT DOMAIN 1872 1878
FT CARRYHYD 594 N-LINKED
FT CARRYHYD 621 N-LINKED
FT CARRYHYD 672 N-LINKED
FT CARRYHYD 747 N-LINKED
FT CARRYHYD 823 N-LINKED
FT CARRYHYD 1172 N-LINKED
FT CARRYHYD 1522 N-LINKED
FT CARRYHYD 1658 N-LINKED
SQ SEQUENCE 1905 AA; 212518 MW; B8E223FA8B9AE13C CRC64;

Query Match 22 2%; Score 61; DB 1; Length 1905;
Best Local Similarity 28 3%; Pred. No. 33; Mismatches 13; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

DR RESULT 8
ID GBL1_MAIZE STANDARD; PRT; 573 AA.
ID GBL1_MAIZE STANDARD; PRT; 573 AA.
AC P15590;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DR GLOBULIN-1 S ALLELE PRECURSOR (GBL1-S) (7S-LIKE).
DR GBL1
OS Zea mays (Maize);
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. INBRED LINE VA26;
RA Belanger F.C., Kriz A.L.;
RT "Molecular characterization of the major maize embryo globulin encoded by the Gb1 gene";
RL Plant Physiol. 91:636-643(1989).
RN [2]
RP SEQUENCE OF 87-100.
RX MEDLINE=93374022; PubMed=2775172;
RA Kriz A.L.;
RT "Characterization of embryo globulins encoded by the maize Gb genes";
RT Biocem. Genet. 27:239-251(1989).
RC -!- MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
CC -!- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GBL1 ALLELES HAVE THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL PROTEINS, RESPECTIVELY.
CC -!- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICILIN, CONGLYCIIN, ETC.).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
CC
DR EMBL; M2845; AAA3467.1; -
DR HSSP; P50477; ICALW.
DR MATEDB; 30181; -
DR INTERPRO; IPR001113; -
DR PFAM; PF00546; Seedstore_7s; 1.
DR KW Seed storage protein; Signal; OR 21 (POTENTIAL).
FT SIGNAL 18
FT PROPEP 19 86
FT CHAIN 87 573 GLOBULIN-1 S ALLELE.
FT CARRYHYD 349 349 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 573 AA; 65029 MW; 525ED100A062976 CRC64;

Query Match 21 6%; Score 59.5; DB 1; Length 573;
Best Local Similarity 38 9%; Pred. No. 16; Mismatches 14; Conservative 7; Mismatches 12; Indels 3; Gaps 2;

DR RESULT 9
ID DPOA TRYBB STANDARD; PRT; 3339 AA.
ID DPOA TRYBB STANDARD; PRT; 3339 AA.
AC P2777;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DR DNA POLYMERASE ALPHA CATALYTIC SUBUNIT (EC 2.7.7.7) (DNA POLYMERASE

DE I).

OS Trypanosoma brucei brucei.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.

RN [1]

RP SPOUSENCE FROM N.A.

RC STRAINS5427;

RX MEDLINE=92033600; PubMed=1754381;

RA Leegwater, P.A.J.; Strating, M.S.; Murphy, N.B.; Kooy, R.F.,

RA Van der Vliet, P.C.; Overduine, J.P.;

RT "the Trypanosoma brucei DNA polymerase alpha core subunit gene is developmentally regulated and linked to a constitutively expressed open reading frame.";

RT NUCLEIC ACIDS Res. 19:6441-6447(1991).

CC -!- FUNCTION: POLYMERASE ALPHA IN A COMPLEX WITH DNA PRIMASE IS A REPLICATIVE POLYMERASE.

CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE + DNA(N).

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERSSES: ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR DIFFERENT REACTIONS OF DNA SYNTHESIS.

CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X60951; CAA43287.1; -.

CC PIR: S20052; S20052.

CC INTERPRO: IPR002064; -.

CC PFAM: PF00136; DNA_POL_B; 1.

CC PRINTS: PRO0106; DNAPOLB.

DR PROSITE: PS00116; DNA_POLMERASe_B; 1.

KW PROTEIN: DNA_POLMERASe_B; 1.

KW DNA-binding; Nuclear protein.

SEQUENCE 1339 AA; 151611 MW; 9FF159412F2B7FBA CRC64;

SQ

Query Match 21.5%; Score 59; DB 1; Length 1339;

Best Local Similarity 32.3%; Pred. No. 39; Matches 10; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

Oy 17 EWDTRGQEKODQCESCKSQYGEKEQDQQRHR 47

Db 3 DWVSCRSEEQRCERGQSTFDESEBEEQWR 33

RESULT 10

FBL4_CRIGR FBL4_CRIGR STANDARD; PRT; 443 AA.

ID FBL4_CRIGR STANDARD; PRT; 443 AA.

AC 055058; -

DT 05-OCT-2000 (Rel. 40, Created)

DT 01-OCT-2000 (Rel. 40, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR MATRIX PROTEIN 2 PRECURSOR (FIBULIN-4) (FBL4) (H411 PROTEIN).

GN EFEMP2 OR FBLN4.

OS Cricetulus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Rodentia; Sciurognathii; Muridae; Cricetinae; Cricetulus.

[1]

RP SEQUENCE FROM N.A.

RC TISSUE=OVARY;

RA Heine, H.; Delude, R.L.; Monks, B.; Golenbock, D.T.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

CC -!- SUBCELLULAR LOCATION: SECRETORY.

CC -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AF046870; AAC03101.1; -.

DR HSSP: P00736; IAPQ.

DR INTERPRO: IPR000152; -.

DR INTERPRO: IPR000561; -.

DR PROSITE; PS00008; EGF; 4.

DR PROSITE; PS00010; ASX_HYDROXYL; 4.

DR PROSITE; PS00022; EGF_1; FALSE_NEG.

DR PROSITE; PS01186; EGF_2; 4.

DR PROSITE; PS01187; EGF_CA; 6.

KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.

FT SIGNAL 1 25

FT CHAIN 26 443

FT DOMAIN 36 81

FT DOMAIN 123 163

FT DOMAIN 164 202

FT DOMAIN 203 242

FT DOMAIN 243 282

FT DOMAIN 283 328

FT DISULFID 127 140

FT DISULFID 134 149

FT DISULFID 151 162

FT DISULFID 168 177

FT DISULFID 173 186

FT DISULFID 188 201

FT DISULFID 207 217

FT DISULFID 213 226

FT DISULFID 228 241

FT DISULFID 247 258

FT DISULFID 254 267

FT DISULFID 269 281

FT DISULFID 287 300

FT DISULFID 294 309

FT DISULFID 315 327

FT CARBOHYD 198 198

FT CARBOHYD 394 394

FT SEQUENCE 443 AA; 49432 MW; 0BCFED7323D9E5F CRC64;

Query Match 21.3%; Score 58.5; DB 1; Length 443;

Best Local Similarity 30.0%; Pred. No. 16; Matches 12; Conservative 7; Mismatches 16; Indels 5; Gaps 2;

Oy 2 DDDPPKRYECCRREWDYRQG--KEQQQC--EESCKSQ 36

Db 29 DSEEPDSYTBCTDGYEWDAQSOHQCRDVNECLTIPEACKE 68

RESULT 11

MIG1_KLUMA MIG1_KLUMA STANDARD; PRT; 543 AA.

ID MIG1_KLUMA MIG1_KLUMA STANDARD; PRT; 543 AA.

AC P52288; -

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE REGULATORY PROTEIN MIG1.

GN

OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis);

OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;

OC Saccharomyctaceae; Kluyveromyces.

[1]

RN SEQUENCE FROM N.A.

RP STRAIN=ATCC 1424;

RA Cassart, J.; Ostling, J.; Ronne, H.; Delcour, J.; Vandenhoute, J.;

CC

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: INVOLVED IN GLUCOSE REPRESSION OF GLUCOSE METABOLISM

CC GENES

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- SIMILARITY: TO THE MAMMALIAN EGR (EARLY GROWTH RESPONSE) PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: Z50018; CAA90321.1; -.

DR HSSP: P0846; I11L.

DR INTERPRO: IPR000822; -.

PFAM: PF0096; zf-C2H2; 2.

PRINTS: PR00048; ZINC_FINGER.

DR PR00048; ZINC_FINGER.

KW TRANSCRIPTION REGULATION; REPRESSOR; DNA-BINDING; NUCLEAR PROTEIN;

ZINC-FINGER; METAL-BINDING; REPEAT; CARBOHYDRATE METABOLISM.

FT ZN_FING 32 54 C2H2-TYPE.

FT DOMAIN 1.1 131 POLY-GLN.

FT DOMAIN 4.99 504 POLY-SER.

SQ SEQUENCE 543 AA; 59315 MW: DFC01BD92144622F CRC64;

Query Match 21.3%; Score 58.5; DB 1; Length 1085; Best Local Similarity 31.7%; Pred. No. 36; Matches 13; Conservative 14; Mismatches 13; Indels 1; Gaps 1; Query Match 21.3%; Score 58.5; DB 1; Length 1085; Best Local Similarity 31.7%; Pred. No. 36; Matches 13; Conservative 14; Mismatches 13; Indels 1; Gaps 1; DR 01-JAN-1990 (Rel. 13, Last sequence update) DT 01-AUG-1992 (Rel. 23, Last annotation update)

DE LEGUMIN B (FRAGMENT).

GN LEGB.

OS Pisum sativum (Garden pea).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Fabales; Fabaceae; Papilionoideae; Pisum.

RN [1]

RP Domoney C., Barker D., Casey R.;

RA "The complete deduced amino acid sequences of legumin beta-polypeptides from different genetic loci in Pisum.";

RT Plant Mol. Biol. 7:467-474 (1986).

CC -!- FUNCTION: THIS PROTEIN FOUND IN THE SEEDS OF MANY LEGUMINOUS & NONLEGUMINOUS PLANTS IS THE SOURCE OF SULFUR-CONTAINING AMINO ACIDS IN SEED MEALS.

CC -!- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A DISULFIDE BOND.

CC -!- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEINS (GLOBULINS) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M16890; AAA31678.1; -.

DR PIR: S04321; S04321.

DR INTERPRO: IPR000459; -.

PFAM: PF00190; Seedstore_11s; 2.

DR PROSITE: PS00305; 11S_SEED_STORAGE; 1.

DR KW Seed storage protein; Multigene family.

FT NON_TER 1

FT CHAIN <1 161 ALPHA CHAIN (ACIDIC).

FT CHAIN 162 338 BETA CHAIN (BASIC).

FT DISLFID 168 168 INTERCHAIN (ALPHA-BETA) (POTENTIAL).

SQ SEQUENCE 338 AA; 38989 MW: 752CFCD33686AE0 CRC64;

Query Match 21.1%; Score 58; DB 1; Length 338; Best Local Similarity 29.6%; Pred. No. 14; Matches 16; Conservative 12; Mismatches 12; Indels 14; Gaps 3; DR EMBL: Z66535; CAA91432.1; -.

CC Hypothetical protein.

CC DOMAIN 33 46 POLY-LYS.

FT DOMAIN 184 194 POLY-LYS.

FT DOMAIN 695 698 POLY-GLU.

FT DOMAIN 714 717 POLY-GLU.

FT DOMAIN 718 721 POLY-ARG.

RESULT 13

LEGB_PEA

ID LEGB_PEA STANDARD; PRT; 338 AA.

AC P14394; -.

DR 01-JAN-1990 (Rel. 13, Last sequence update)

DE LEGUMIN B (FRAGMENT).

GN LEGB.

OS Pisum sativum (Garden pea).

OC Magnoliophyta; eudicots; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Pisum.

RN [1]

RP Domoney C., Barker D., Casey R.;

RA "The complete deduced amino acid sequences of legumin beta-polypeptides from different genetic loci in Pisum.";

RT Plant Mol. Biol. 7:467-474 (1986).

CC -!- FUNCTION: THIS PROTEIN FOUND IN THE SEEDS OF MANY LEGUMINOUS & NONLEGUMINOUS PLANTS IS THE SOURCE OF SULFUR-CONTAINING AMINO ACIDS IN SEED MEALS.

CC -!- SUBUNIT: HEXAMER; EACH SUBUNIT IS COMPOSED OF AN ACIDIC AND A BASIC CHAIN DERIVED FROM A SINGLE PRECURSOR AND LINKED BY A DISULFIDE BOND.

CC -!- SIMILARITY: BELONGS TO THE 11S SEED STORAGE PROTEINS (GLOBULINS) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M16890; AAA31678.1; -.

DR PIR: S04321; S04321.

DR INTERPRO: IPR000459; -.

PFAM: PF00190; Seedstore_11s; 2.

DR PROSITE: PS00305; 11S_SEED_STORAGE; 1.

DR KW Seed storage protein; Multigene family.

FT NON_TER 1

FT CHAIN <1 161 ALPHA CHAIN (ACIDIC).

FT CHAIN 162 338 BETA CHAIN (BASIC).

FT DISLFID 168 168 INTERCHAIN (ALPHA-BETA) (POTENTIAL).

SQ SEQUENCE 338 AA; 38989 MW: 752CFCD33686AE0 CRC64;

Query Match 21.1%; Score 58; DB 1; Length 338; Best Local Similarity 29.6%; Pred. No. 14; Matches 16; Conservative 12; Mismatches 12; Indels 14; Gaps 3; DR 107 DEDEPPRSVETRK--WKKHTAEKKRESHGQEEE- BELEKEEEEEEIQORH 155

RESULT 14

FBLA4_HUMAN

ID FBL4_HUMAN STANDARD; PRT; 443 AA.
AC 09567; 075967; PRT;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR MATRIX PROTEIN 2 PRECURSOR
(FIBULIN-4) (FBLN-4) (UPH1 PROTEIN).
DE EFEMP2 OR FBLNA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.
RN [1] SEQUENCE FROM N.A.
RT TISSUE=MELOMOMA;
RX MEDLINE=20068041; PubMed=10601734;
RA Giltay R.; Timpl R.; Kostka G.;
RT "Sequence, recombinant expression and tissue localization of two novel extracellular matrix proteins, fibulin-3 and fibulin-4";
RL Matrix Biol. 18:469-480(1999).
RN [2] SEQUENCE FROM N.A.
RA Submitted (SEP-1998) to the EMBL/GenBank/DDJB databases.
RN [3] SEQUENCE FROM N.A.
RA Katsanis N., Venable S., Smith J.R., Lupski J.R.:
RT "Isolation of a paralog of the Doyne honeycomb retinal dystrophy gene from the multiple retinopathy critical region on 11q13.";
RL Hum. Genet. 105:66-72(2000).
CC !- SUBCELLULAR LOCATION: SECRETED.
CC !- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
EMBL: AJ112819; CAAI0791_2; -;
DR EMBL; AF093119; AAC22108_1; -;
DR EMBL; AF109121; AAC65188_1; -;
DR HSSP; P3555; IEMO.
DR INTERPRO; IPR000152; -;
DR PRINS; PRO0907; THRBOMODULIN.
DR PROSITE; PS000010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01187; EGF_CA; 6.
DR KW repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 443 EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR MATRIX PROTEIN 2.
FT DOMAIN 36 81 EGF-LIKE 1, DIVERGENT.
FT DOMAIN 123 163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 164 242 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 203 242 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 243 282 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 283 328 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DISULFID 127 140 BY SIMILARITY.
FT DISULFID 134 149 BY SIMILARITY.
FT DISULFID 151 162 BY SIMILARITY.
FT DISULFID 168 177 BY SIMILARITY.
FT DISULFID 173 186 BY SIMILARITY.
FT DISULFID 188 201 BY SIMILARITY.
FT DISULFID 207 217 BY SIMILARITY.
FT DISULFID 213 226 BY SIMILARITY.

DT DISULFID 228 241 BY SIMILARITY.
FT DISULFID 247 258 BY SIMILARITY.
FT DISULFID 254 267 BY SIMILARITY.
FT DISULFID 269 281 BY SIMILARITY.
FT DISULFID 287 300 BY SIMILARITY.
FT DISULFID 294 309 BY SIMILARITY.
FT DISULFID 315 327 BY SIMILARITY.
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 5 5 A -> T (IN REF. 1).
FT CONFLICT 44 51 EWDPDSQH -> TQFAN (IN REF. 2).
FT CONFLICT 103 111 AQHPNIPPP -> VNTQPLPT (IN REF. 2).
FT CONFLICT 294 294 C -> W (IN REF. 2).
FT CONFLICT 354 356 RSV -> AER (IN REF. 2).
FT CONFLICT 355 355 S -> R (IN REF. 3).
SQ SEQUENCE 443 AA; 49391 MW; 9E9ACC2393780D3B8 CRC64;

RESULT 15 Query Match 20.9%; Score 57.5; DB 1; Length 443;
FBL4_MOUSE STANDARD; PRT; 443 AA.
ID FBL4_MOUSE
AC 09WV19;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR MATRIX PROTEIN 2 PRECURSOR
(FIBULIN-4) (FBLN4) (MBP1).
DE EFEMP2 OR FBLNA OR MBP1.
OS Mus musculus (Mouse).
OC Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1] SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J;
RX MEDLINE=9308389; PubMed=10380882;
RA Gallagher W.M.; Argentini M.; Sierra V.; Bracco L.; Debusseche L..
RA Conseiller E.;
RP "MBP1: a novel mutant p53-specific protein partner with oncogenic properties";
RT Oncogene 18:3608-3616(1999).
CC !- SUBCELLULAR LOCATION: SECRETED.
CC !- SUBUNIT: BIND PREFERENTIALLY TO P53 MUTANTS.
CC !- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
EMBL: AF104223; AAC45219_1; -;
DR INTERPRO; IPR000152; -;
DR PRINS; PRO0907; THRBOMODULIN.
DR PROSITE; PS000010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 4.
DR PROSITE; PS01187; EGF_CA; 6.
KW Repeat; EGF-like domain; Calcium-binding; Glycoprotein; Signal.

FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 443 EGF-CONTAINING FIBULIN-LIKE EXTRACELLULAR
 MATRIX PROTEIN 2.
 FT DOMAIN 36 81 EGF-LIKE 1, DIVERGENT.
 FT DOMAIN 123 163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 164 202 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 203 242 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 243 282 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 283 328 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 127 140 BY SIMILARITY.
 FT DISULFID 134 149 BY SIMILARITY.
 FT DISULFID 151 162 BY SIMILARITY.
 FT DISULFID 168 177 BY SIMILARITY.
 FT DISULFID 173 186 BY SIMILARITY.
 FT DISULFID 188 201 BY SIMILARITY.
 FT DISULFID 207 217 BY SIMILARITY.
 FT DISULFID 213 226 BY SIMILARITY.
 FT DISULFID 228 241 BY SIMILARITY.
 FT DISULFID 247 258 BY SIMILARITY.
 FT DISULFID 254 267 BY SIMILARITY.
 FT DISULFID 269 281 BY SIMILARITY.
 FT DISULFID 287 300 BY SIMILARITY.
 FT DISULFID 294 309 BY SIMILARITY.
 FT DISULFID 315 327 BY SIMILARITY.
 FT CARBOHYD 198 394 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 394 443 AA; 49425 MW; 4969C0328A23DD88 CRC64;
 SQ SEQUENCE

Query Match 20 9%; Score 57.5; DB 1; Length 443;
 Best Local Similarity 30.0%; Pred. No. 20;
 Matches 12; Conservative 7; Mismatches 16; Indels 5; Gaps 2;

QY 2 DDDPKRYEDCRRREWDTRGQ--KEQQQC--EESCKSQ 36
 Db 29 DPEREDPSYECTDGWEWADSOHCRRDVNECLTPEACKGE 68

Search completed: March 1, 2001, 16:16:55
 Job time: 435 sec